

ANALISIS KURASI DATA PADA REPOSITORY LEMBAGA PENELITIAN DUNIA

Abdurrahman Prasetyadi^{1*}; Atam Ependi²; D.W. Ari Nugroho³; Didik Prata Wijaya⁴

^{1,2,3,4} Pusat Data dan Dokumentasi Ilmiah - Lembaga Ilmu Pengetahuan Indonesia

*Korespondensi: abdurrahmanpras@gmail.com

ABSTRAK

Setiap lembaga penelitian memiliki cara dan kebijakan yang berbeda terkait kurasi data penelitiannya. Kajian ini bertujuan untuk: (1) mengetahui proses kurasi data yang ideal di lembaga penelitian dunia; dan (2) menganalisis perbandingan dokumentasi data pada sistem repository lembaga penelitian dunia. Metode yang digunakan adalah deskriptif kualitatif dengan tahapan studi literatur, wawancara, mensurvey sistem repository, dan menginterpretasikan data. Hasil dari penelitian ini menunjukkan proses kurasi data yang ideal di lembaga penelitian yaitu kurator terlibat sejak proses akuisisi sampel data, mengumpulkan data mentah, menilai kualitas data, memproses data mentah menjadi data turunan, mendokumentasikan dan melengkapi metadata, mengatur akses kontrol hingga menemukan kembali data. Selanjutnya, perbandingan dokumentasi data dapat dilihat pada proses penentuan direktori/*dataverse*, pelengkapan metadata, penggunaan bibliografi terkontrol, dan sebagainya. CIFOR dianggap sebagai lembaga penelitian dengan proses kurasi dan pendokumentasian data yang paling ideal.

Kata kunci: data penelitian, kurasi data, lembaga penelitian dunia, repository, deposit

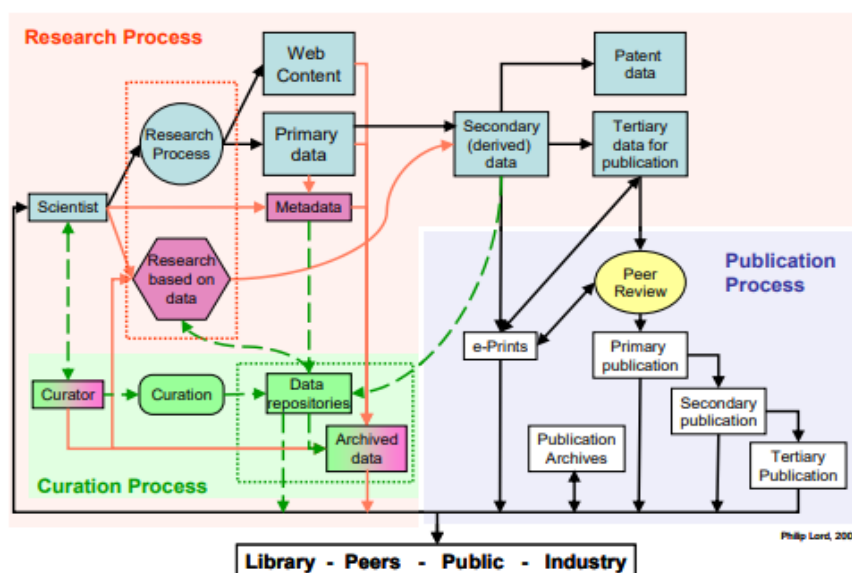
1. PENDAHULUAN

Saat ini, lembaga penelitian dunia sudah mengembangkan berbagai macam platform repository khusus agar peneliti dapat mendepositkan, mengelola, dan memelihara data penelitiannya. Bahkan, di berbagai negara maju, pemerintah memiliki tanggung jawab untuk mengelola data penelitian ilmiah masyarakatnya, seperti di Inggris dengan UK Data Service, Belanda dengan DANS, dan Australia dengan ANDS. Adapun manfaat dari mendepositkan data di repository, yaitu dapat meningkatkan visibilitas penelitian, kolaborasi ilmiah, dan akses oleh masyarakat. Ketika mendepositkan datanya di repository, seorang peneliti sekaligus memiliki peran sebagai pengkurasi data (kurator) yang melengkapi dan memvalidasi datanya apakah layak untuk dipublikasikan.

Makna dari kurasi adalah pemilihan, pelestarian, pemeliharaan, pengumpulan, dan pengarsipan aset penelitian (Scime 2009). Sementara, menurut Miller (2014) kurasi data adalah organisasi dan integrasi data yang dikumpulkan dari berbagai sumber. Kurasi melibatkan anotasi, publikasi, dan penyajian data sedemikian rupa sehingga nilai data dipertahankan dari waktu ke waktu, dan data tetap tersedia untuk digunakan kembali dan disimpan.

Lebih lanjut lagi, istilah kurasi semakin banyak digunakan untuk tindakan yang diperlukan dalam menjaga data penelitian digital dan bahan digital lainnya atas seluruh siklus hidup data dari waktu ke waktu untuk generasi sekarang dan masa depan pengguna (Giaretta 2005 dalam Bakry 2016). Data penelitian sendiri merupakan berbagai fakta dan angka yang dapat dijadikan bahan untuk menyusun suatu informasi ilmiah (Arikunto 2002).

Beberapa repositori lembaga penelitian dunia di Indonesia misalnya yang sudah mempublikasikan data penelitian, yaitu LIPI dan CIFOR, telah menunjuk kurator dari ketua kelompok penelitian dan manajer data untuk proses kurasi. Pada umumnya jenis data yang dikurasi antara lain seperti teks, rekaman video, foto, dan audio penelitian. Adapun sumber data diperoleh dari hasil observasi, catatan laboratorium, dan model komputasi.



Gambar 1 Model Proses Kurasi Data Ideal oleh Lord (2003)

Gambar 1 di atas menjelaskan proses kurasi yang idealnya perlu dilakukan oleh peneliti atau kurator di suatu lembaga penelitian. Prosesnya diawali dari data primer yang diidentifikasi, pemberian deskripsi metadata, pengarsipan *datafile*, serta verifikasi dan penyajian data. Namun permasalahannya, bagi banyak lembaga penelitian tidak diketahui secara pasti apakah proses ideal yang disebutkan di atas sudah dilakukan atau tidak.

Berdasarkan fenomena di atas, perlu dilakukan suatu kajian yang tepat terkait analisis kurasi data pada repositori lembaga penelitian untuk mengetahui cara kurasi yang beragam dan ideal. Manfaat dari kajian ini diharapkan dapat memberikan masukan untuk para pemangku kebijakan di Pusat Data dan Dokumentasi Ilmiah (PDDI) LIPI sebagai pengembang Repositori Ilmiah Nasional (RIN).

Tujuan dari kegiatan ini yaitu: (1) mengetahui proses kurasi data yang ideal di lembaga penelitian dunia; dan (2) menganalisis perbandingan dokumentasi data pada sistem repositori lembaga penelitian dunia. Beberapa kajian terkait repositori yang peneliti jadikan referensi diantaranya dijelaskan dalam Tabel 1 berikut.

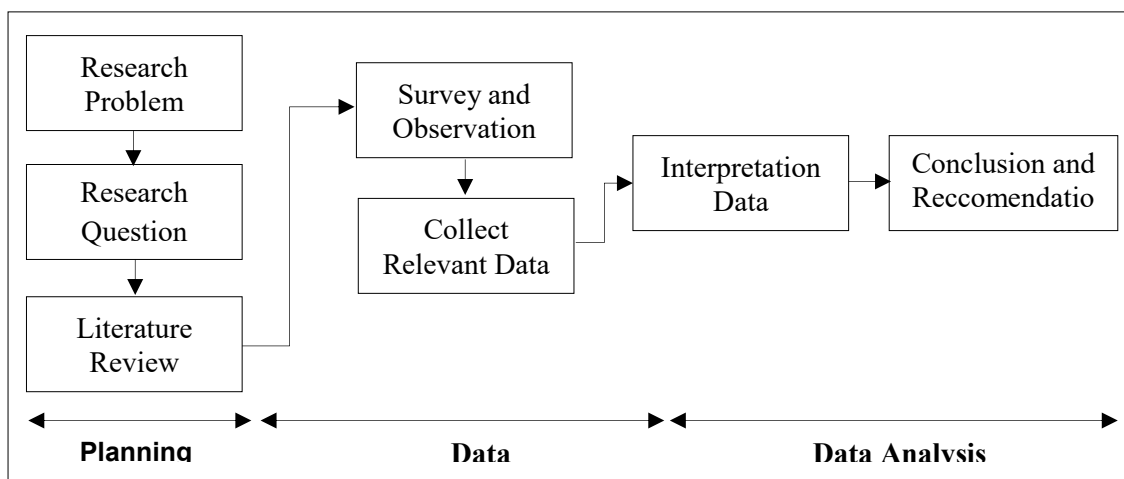
Tabel 1 Penelitian Terdahulu

Peneliti	Judul	Deskripsi Penelitian
Anjas Alifah Bakry (2016)	Kurasi Digital untuk Pengarsipan Digital	Kurasi digital memiliki implikasi untuk berbagai sektor yang berbeda karena mereka bergerak dari kertas ke lingkungan digital. Bagi masyarakat dan individu, bisa dikatakan bahwa pengetahuan digital jika ingin bermanfaat dan bisa digunakan harus terus diperbarui, dipelihara, dan diakses. Upaya yang signifikan perlu dimasukkan ke dalam pengembangan infrastruktur informasi secara terus-menerus untuk bahan digital.
Michael Witt (2008)	Institutional Repositories and Research Data Curation in a Distributed Environment	Secara garis besar, tidak adanya kerangka kerja untuk pengorganisasian dan pelestarian data penelitian yang tersedia untuk jangka panjang telah membuat <i>dataset</i> berharga akan menjadi hilang atau dibuang. Pendekatan dari Pusat Kurasi Data dari Perpustakaan Universitas Purdue telah mengintegrasikan para pustakawan dan prinsip-prinsip perpustakaan dengan ilmu komputer untuk mengatasi tantangan mengelola koleksi data penelitian dan untuk belajar bagaimana mendukung penelitian interdisipliner melalui kurasi data.
G. Sayeed Choudhury (2008)	Case Study in Data Curation at Johns Hopkins University	Johns Hopkins University telah mengembangkan <i>institutional repository</i> (IR) sebagai komponen arsitektur perpustakaan digital secara keseluruhan, yang akan menekankan kepada pelestarian data jangka panjang. Pada dasarnya, IR tersebut dikembangkan sebagai " <i>gateway</i> " ke arsip digital untuk mendukung kurasi data sebagai bagian dari infrastruktur <i>cyber</i> yang akan menampilkan <i>open data</i> .

2. METODE

Metode yang digunakan dalam penelitian ini yaitu metode kualitatif dengan pendekatan deskriptif, yaitu penelitian yang memberi gambaran secara cermat mengenai individu atau kelompok tertentu tentang keadaan dan gejala yang terjadi (Koentjaraningrat 1981). Tahapannya terdiri dari:

- a) melakukan studi literatur penelitian sebelumnya mengenai kurasi data;
- b) melakukan wawancara terhadap kurator di salah satu lembaga penelitian dunia (CIFOR);
- c) melakukan survei kurasi data pada sistem repositori lembaga penelitian dunia; dan
- d) menganalisis dan menginterpretasikan keberagaman kurasi data di lembaga penelitian dunia.



Gambar 2 Tahapan Penelitian yang Dilakukan (diolah dari berbagai sumber)

Pada Gambar 2, dijelaskan bahwa proses penelitian ini diawali dengan adanya masalah penelitian yang kemudian dibuat menjadi pertanyaan penelitian. Kemudian pada tahap perencanaan dilakukan juga studi literatur dari penelitian-penelitian sebelumnya. Tahap selanjutnya yaitu pengumpulan data dengan mensurvei, mengobservasi, dan menyortir data-data yang relevan. Terakhir, hasil pengumpulan data akan diinterpretasikan secara deskriptif, serta dibuat kesimpulan dan saran sebagai rekomendasi ilmiah.

Tabel 2 Objek Penelitian

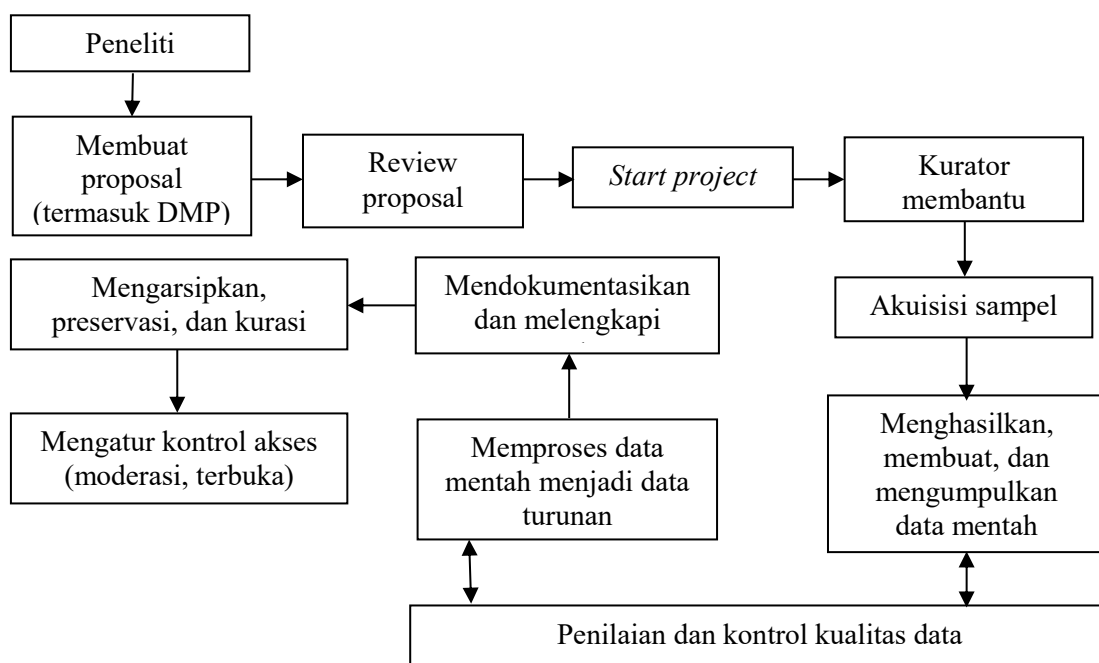
No.	Nama Repositori	Lembaga Penelitian	Asal Negara	Alamat Situs Repositori
1	Meldata	ICARDA	Lebanon	https://data.mel.cgiar.org/
2	Data Archiving and Networked Services (DANS)	Netherlands Organisation for Scientific Research (NWO)	Belanda	https://dataverse.nl/
3	CIFOR	CIFOR	Indonesia	https://data.cifor.org
4	AUSSDA	AUSSDA	Austria	https://data.aussda.at
5	CIRAD	Cirad	Perancis	https://dataverse.cirad.fr
6	DataSpace	DataSpace	Hongkong	https://dataspace.ust.hk
7	Gottingen Research Online/Data	Gottingen Research Online	Jerman	https://data.goettingen-research-online.de/

Penelitian ini dilakukan di Kepustakaan Kawasan PDDI Bandung di Komplek LIPI Jalan Sangkuriang, Bandung. Waktu penelitian dilaksanakan selama 5 bulan sejak bulan April sampai dengan September tahun 2019. Daftar objek penelitian ini dapat dilihat pada Tabel 2.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Penelitian

Lembaga riset yang diamati secara mendetail yaitu CIFOR (*Center for International Forestry Research*). Hasil wawancara dengan kurator (manajer data) CIFOR Bogor, menjelaskan bahwa proses kurasi data sudah diterapkan dalam perencanaan manajemen data (DMP) pada proposal penelitian. Dalam hal ini kurator sudah dilibatkan untuk membantu peneliti sejak mengakuisisi sampel, membuat dan mengumpulkan data mentah, hingga menemukan data kembali. Berikut interpretasi peneliti terkait proses lengkap kurasi data di CIFOR, dapat dilihat pada Gambar 3.



Gambar 3 Proses Kurasi Data di CIFOR (Hasil Wawancara)

Proses kurasi data seperti gambar di atas diawali ketika peneliti membuat usulan penelitiannya dalam bentuk proposal, di mana manajer data (kurator) akan membantu proses akuisisi sampel dan data mentahnya sekaligus menilai dan mengontrol kualitas datanya menjadi data turunan. Selanjutnya ketika mendokumentasikan *dataset*, kurator CIFOR juga membuat metadata di dalam *datafile* yang mendeskripsikan *header*, jenis data, tanggal

dokumentasi, dan sebagainya. Setelah semuanya selesai, data primer akan diarsipkan dan dipreservasi dengan terlebih dahulu diatur akses kontrol dan penyebarannya.

3.2 Pembahasan

Hasil survei kepada 7 (tujuh) sistem repositori lembaga penelitian dunia yang seluruhnya berbasis *dataverse* menunjukkan perbandingan terkait dokumentasi data penelitiannya. Seperti penentuan direktori/*dataverse*, pelengkapan metadata, penggunaan bibliografi terkontrol, dan pengaturan publikasi data yang berbeda. Beberapa tabel berikut merupakan hasil survei dari kajian ini.

Tabel 3 Penentuan Direktori/*Dataverse*

Sistem Repositori	Direktori/<i>Dataverse</i>	<i>Dataverse</i> dan Sub <i>Dataverse</i>
<i>Meldata</i>	Berdasarkan subjek penelitian dari suatu pusat penelitian	<i>Dataverse: Big data and ICT dataverse</i> <i>Sub dataverse: Tidak ada, langsung dataset</i>
<i>DANS</i>	Berdasarkan proyek penelitian	<i>Dataverse: DANS projects dataverse</i> <i>Sub dataverse: Tidak ada, langsung dataset</i>
<i>CIFOR</i>	Berdasarkan program kolaborasi CIFOR dengan lembaga lainnya	<i>Dataverse: The Sustainable Wetlands Adaptation and Mitigation Program (SWAMP)</i> <i>Sub dataverse: Database of tropical wetlands carbon survey: Spatial Data</i>
<i>AUSSDA</i>	Berdasarkan universitas/institut, hasil survei, dan jurnal	<i>Dataverse: SORA dataverse (SORA - Institute for Social Research and Consulting)</i> <i>Sub dataverse: Tidak ada, langsung dataset</i>
<i>CIRAD</i>	Berdasarkan unit penelitian kolaborasi	<i>Dataverse: UMR System</i> <i>Sub dataverse: Tropical Agroforestry Systems</i>
<i>DataSpace</i>	Berdasarkan peneliti	<i>Dataverse: Lee-Campbell research group dataverse</i> <i>Sub dataverse: Tidak ada, langsung dataset</i>
<i>Gottingen Research Online/Data</i>	Berdasarkan peneliti dan/atau departemen penelitian	<i>Dataverse: Sven Bingert dataverse</i> <i>Sub dataverse: Tidak ada, langsung dataset</i>

Dari Tabel 3, terlihat bahwa sebagian besar lembaga penelitian tidak memiliki *sub dataverse* tambahan, hanya CIRAD dan CIFOR yang memilikinya. Hal tersebut dapat mengindikasikan bahwa data-data penelitian tersebut terintegrasi ke dalam satu *dataverse* yang bisa memudahkan pengelolaannya.

Tabel 4 Pelengkapan Metadata

Sistem Repositori	Jenis Metadata yang Ditambahkan	Keterangan
<i>Meldata</i>	<i>Geospatial metadata, social science and humanities metadata, dan astronomy and astrophysics metadata</i>	Metadata geospasial untuk deskripsi cakupan geografis, sosial untuk deskripsi jumlah sampel penelitian dan teknik pengumpulan data, dan astronomi untuk deskripsi resolusi spasial
<i>DANS</i>	Tidak ada yang ditambahkan	Untuk metadata sitasi
<i>CIFOR</i>	<i>Geospatial metadata, Administrative Information Metadata, dan overview of the research</i>	Metadata geospasial untuk deskripsi cakupan geografis, informasi administrasi untuk deskripsi tipe fisik media dan lokasi media, dan tinjauan untuk deskripsi area fokus penelitian dan sumber dana
<i>AUSSDA</i>	<i>Geospatial metadata, dan social science and humanities metadata</i>	Metadata geospasial untuk deskripsi cakupan geografis, sosial untuk deskripsi jumlah sampel penelitian dan teknik pengumpulan data
<i>CIRAD</i>	<i>Geospatial metadata dan life science metadata</i>	Metadata ilmu hayati digunakan untuk deskripsi nama organisme dan cara pengukuran
<i>DataSpace</i>	<i>Geospatial metadata, dan social science and humanities metadata</i>	Metadata geospasial untuk deskripsi cakupan geografis, sosial untuk deskripsi jumlah sampel penelitian dan teknik pengumpulan data
<i>Gottingen Research Online/Data</i>	<i>Geospatial metadata, dan social science and humanities metadata</i>	Metadata geospasial untuk deskripsi cakupan geografis, sosial untuk deskripsi jumlah sampel penelitian dan teknik pengumpulan data

Tabel 4 menunjukkan bahwa sebagian besar lembaga penelitian menambahkan metadata *geospatial metadata*, dan *social science and humanities metadata*. Hanya satu lembaga (*DANS*) yang menggunakan metadata sitasi secara *default (citation metadata)*.

Tabel 5 Penggunaan Bibliografi Terkontrol

Sistem Repositori	Penggunaan Bibliografi Terkontrol	Keterangan
<i>Meldata</i>	Menggunakan	http://aims.fao.org/vest-registry/vocabularies/agrovoc
<i>CIFOR</i>	Menggunakan	http://browser.agrisemantics.org/gacs/en/page/C7406

Sistem Repositori	Penggunaan Bibliografi Terkontrol	Keterangan
<i>AUSSDA</i>	Menggunakan	https://elsst.ukdataservice.ac.uk/
<i>CIRAD</i>	Menggunakan	http://aims.fao.org/vest-registry/vocabularies/agrovoc
<i>DataSpace</i>	Tidak menggunakan	-
<i>Gottingen Research Online/Data</i>	Tidak menggunakan	-

Tabel 5 menunjukkan empat lembaga (Meldata, CIFOR, AUSSDA, CIRAD) menggunakan bibliografi terkontrol dalam penentuan kata kunci datanya, dan tiga lembaga tidak menggunakannya (DANS, DataSpace, Gottingen Research Online/Data).

Tabel 6 Beberapa Jenis Berkas Digital yang Disimpan

Sistem Repositori	Berkas digital	Keterangan
<i>Meldata</i>	<i>TIFF Image, CSV (Comma Separated Values), JPEG, eqa, cpq, dbf (unknown)</i>	Menggunakan perangkat lunak khusus (unknown)
<i>DANS</i>	<i>Tar.gz, PNG, PDF, CSV, SPSS SAV (sav), PDF</i>	Banyak berkas teks
<i>CIFOR</i>	<i>Xlsx, zip, XML, JPEG, Tab-Delimited (tab), 7Z Archive (7z), HTML, application/x-endnote-library-archive (ENLX)</i>	Banyak berkas kompresi
<i>AUSSDA</i>	<i>PDF, ZIP Archive (zip), Tab-Delimited (tab), Stata Binary (dta), application/x-sas-syntax (sas), R Syntax (R), application/x-spss-syntax (sps)</i>	Banyak berkas statistik
<i>CIRAD</i>	<i>Xlsx, Comma Separated Values (csv), fas (Unknown), oly (Unknown), text/nlogo (nlogo), Plain Text (txt), application/x-gzip (gzip), XML (xml)</i>	Menggunakan perangkat lunak khusus (unknown)
<i>DataSpace</i>	<i>XML (xml), Adobe PDF (pdf), Tabular Data (tab), PNG Image (png)</i>	Banyak berkas teks
<i>Gottingen Research Online/Data</i>	<i>Xlsx, PDF, PNG, Gzip Archive (gz), Shell Script (sh), R Source Code (R), rkdwn Text (md),</i>	Banyak berkas komputasi

Tabel 6 mengindikasikan bahwa setiap lembaga penelitian menggunakan berkas digital yang bervariasi sesuai dengan kebutuhannya. Beberapa berkas digital itu kurang dapat dikenali (*unknown*) karena diperlukan perangkat lunak khusus untuk membacanya.

4. KESIMPULAN

Hasil dari penelitian ini menunjukkan tidak semua lembaga penelitian melaksanakan proses kurasi data yang ideal. CIFOR merupakan lembaga penelitian yang terlibat sejak proses akuisisi data, mengumpulkan data mentah, menilai kualitas data, memproses data mentah menjadi data turunan, mendokumentasikan dan melengkapi metadata, mengatur akses kontrol hingga menemukan kembali data. Selain itu, perbandingan dokumentasi data menunjukkan proses penentuan direktori/dataverse, pelengkapan metadata, penggunaan bibliografi terkontrol dan jenis berkas digital yang digunakan berbeda diantara lembaga penelitian, tergantung ruang lingkup dan kebijakannya masing-masing.

Berdasarkan hasil survei dan analisis terhadap tujuh lembaga penelitian, diperoleh simpulan bahwa seluruh lembaga penelitian selain CIRAD dan CIFOR tidak memiliki *sub dataverse* tambahan, sebagian besar lembaga penelitian selain DANS menambahkan metadata *geospatial* dan metadata *social science and humanities*. Hasil lain juga menunjukkan bahwa empat lembaga (Meldata, CIFOR, AUSSDA, CIRAD) menggunakan bibliografi terkontrol dalam penentuan kata kunci datanya, dan tiga lembaga tidak menggunakan (DANS, DataSpace, Gottingen Research Online/Data). Terakhir, setiap lembaga penelitian menggunakan berkas digital yang bervariasi sesuai dengan kebutuhannya meskipun beberapa kurang dapat dikenali (*unknown*).

5. UCAPAN TERIMA KASIH

Ucapan terima kasih disampaikan kepada Bapak Hendro Subagyo selaku Kepala PDDI LIPI dan Ibu Noorika Retno Widuri selaku Kepala Bidang Repositori PDDI LIPI yang telah menyetujui dan menganggarkan kajian ini dilaksanakan. Serta terima kasih kepada Ibu Sufit selaku manajer data di CIFOR.

DAFTAR PUSTAKA

- Arikunto, S. 2002. Metodologi Penelitian Suatu Pendekatan Proposal. Jakarta: PT. Rineka Cipta.
- Bakry, Anjas Alifah. 2016. "Kurasi Digital Untuk Pengarsipan Digital." Jurnal Ilmu Perpustakaan dan Informasi, Vol 1, No 1 65-73.
- Koentjaraningrat. 1981. Metode-metode Penelitian Masyarakat. Gramedia.

Miller, Renee J. 2014. "Big Data Curation." December. Accessed August 11, 2019.

<http://comad.in/comad2014/keynotes/BigDataCuration-keynote2-COMAD2014.pdf>

Scime, Erin. 2009. "The Content Strategist as Digital Curator." alistapart. December 8. Accessed

Mei 8, 2019. <http://alistapart.com/article/content-strategist-as-digital-curator/>